Advances in Named Entity Recognition: Exploring State-Of-The- Art Methods

Conference Paper · February 2024

Saja Murtadha Hashim
Karabük University

Kürşat Mustafa Karaoğlan
Karabük University

ADVANCES IN NAME ENTITY RECOGNITION: EXPLORING STATE-OF-THE-ART METHODS

Saja Murtadha HASHIM

Kürşat Mustafa KARAOĞLAN

Department of Computer Engineering, Faculty of Engineering, Karabuk University, Turkey

ORCIDs: {0009-0004-6400-1587, 0000-0001-9830-7622}

ABSTRACT

Named Entity Recognition (NER) is all about deciphering and categorizing named entities in open-domain text. Recently, it has been grabbing considerable interest because of its demonstrated capability to enhance the performance of numerous Natural Language Processing (NLP) applications in various areas, such as translation, detecting colloquial and annoying emails, summarizing specific documents, and interacting with others through responses or discussions, and others, making text comprehension more human-friendly. This review article aims to provide a thorough summary and analysis of recent research papers and developments in NER within NLP. The authors meticulously review and clarify the contributions of critical papers published in the last five years, offering insights into their methodologies and developments. Furthermore, a comparative analysis in tabular form highlights critical aspects such as dataset characteristics, accuracy metrics, models used, and other relevant features in these papers. This paper has delved into the current cutting-edge NER, exploring the latest challenges and limitations faced in this field. Moreover, the authors discuss the tools employed in NER, shedding light on their significance in shaping the landscape of this dynamic and evolving research domain. This comprehensive review is a valuable resource for NLP practitioners, researchers, and enthusiasts, providing a nuanced understanding of the recent trends, contributions, and difficulties in NER.

KEYWORDS: Natural Language Processing, Named Entity Recognition, Entity Extraction, Arabic NER, English NER.

1. INTRODUCTION

Natural Language Processing (NLP) encompasses a range of computational techniques guided by theoretical principles, facilitating automatic analysis, and understanding of human language. It is a digital assistant for researchers, assisting in the automated analysis and comprehension of human language [1]. It also helps unearth valuable insights from text data with minimal computational workload, streamlining the process for greater efficiency and ease [2]. The phrase "named entity" (NE) was initially introduced during the Sixth Message Understanding Conference (MUC-6) [3]. It took place in November 1995 as a task focused on identifying the names of organizations, people, and geographic locations in the text, as well as currency, time, and percentage expressions [4], identifying and classifying named entities such as individuals, organizations, locations, dates, and more is a fundamental task in NLP. Over time, many researchers have conducted different types of research in the NLP field, each focusing on different aspects. Hence, it is crucial to ascertain the current status of NLP research and the fields in which it finds extensive application [5]. In the realm of NLP, a task known as information extraction exists. This task involves extracting of pertinent data related to a specific subject from unstructured texts; within Information Extraction, there is a subtask called Named Entity Recognition (NER), which involves identifying and extracting NEs [6,7]. In today's information-rich world, a vast amount of text is generated daily, encompassing everything from news and scholarly articles to social media and financial reports.

Hidden within this wealth of text are valuable insights, crucial for a wide range of purposes such as retrieving information, organizing knowledge, empowering question-answering systems, machine translation, NER [8,9], creating a knowledge repository [10] including text summarization [11], events or trends detection [12], and web mining which involves the identification of entities within social networks [13]. Typically, NE holds a specific contextual significance within the text, represented by a place, a person's name, an organization, a product brand, other proper nouns, and similar entities. NER's objective is to recognize the boundaries of entities and their respective types within a given sentence [14]. As NER technology continues to evolve, it is essential to delve into the state-of-the-art methods developed to enhance its accuracy, efficiency, and applicability.

The primary objectives of this survey are articulated as follows:

- To conduct an exhaustive examination of the latest developments in NER, thereby offering a comprehensive overview of the field's advancements.
- To elucidate contemporary methodologies, algorithms, and tools that are at the forefront of influencing the landscape of NER within NLP.
- To provide readers with a nuanced comprehension of the current capabilities of NER and its implications across diverse domains, underscoring the significance of NER in the realms of computer science and artificial intelligence.
- To explicate the potential applications of NER, thereby illuminating the various opportunities it presents in different domains.
- To position our article as a valuable scholarly resource for NLP practitioners, researchers, and enthusiasts, with the intention of enhancing their involvement and understanding within the dynamic and continually evolving domain of NER.

This paper organizes its content as follows: Section 1 is reserved for the Introduction section of the paper and provides an overview of the topic, laying the foundation for a comprehensive examination of NER in the context of NLP. Section 2 discusses the use of NER by NLP systems and highlights the advantages and applications of this approach. In In Section 3, the practical implementation of NER is delved into, and the tools provided for language understanding are examined. The broader significance of NER within computer science and AI is discussed in Section 4. Section 5 refines the focus to scrutinize NER in English and Arabic, emphasizing its adaptability across linguistic boundaries. Section 6 conducts a thorough literature review, summarizing recent developments in NER over the past two years. Finally, Section 7 synthesizes the accumulated insights in the Discussion and Conclusion, elucidating implications and proposing potential avenues for future research and development at the intersection of NLP and NER.

2. NLP SYSTEMS LEVERAGING THE ADVANTAGES OF NER APPROACHES

The common NLP tasks that leverage the advantages of NER approaches, encompassing basic tasks, are detailed below and organized into headings.

- **Information Retrieval (IR):** The IR system primarily aims to pinpoint relevant documents on a large scale based on user queries. A study [31] found that a substantial percentage of queries in news databases (67.83%, 83.4%, and 38.8%) included NEs. When integrating a NER system into the IR system, treating each NE as a unique term improved performance over the baseline precision across all recall levels [35].

- **Question Answering (QA):** Its systems are designed to answer user-specific questions precisely. A study found that a significant majority (87.7%) of questions in the CLEF1 2004 and 2005 competitions contained at least one NE. Additionally, another study suggests that integrating an NER system can substantially refine the accomplishment of a QA system, in particular when the answer involves a NE. The authors demonstrated that employing a reliable NER system led to a notable increase in the accuracy of answers [35].

- **Machine Translation (MT):** It faces unique challenges when translating NEs, and specialized approaches. In a study [36], the authors highlight the importance of accurately transliterating NEs to avoid considerable error rates in automatic translation. To enhance MT system performance, they experimented with embedding a NER system as a preliminary processing step. The NER system tagged the text to be translated, and subsequent translation methods specific to NEs were applied. Results indicated that this technique outperforms methods that neglect tagging NEs before translation.

- **Text Clustering (TC):** TC is searching result clustering is an NLP duty that organizes the outcomes from a search engine into groups. For example, when dealing with search results for the query "Michael Jordan," a result clustering system groups documents related to Michael Jordan, the basketball player, in one cluster and those relevant to the Berkeley professor in another. In a study [37], the authors improved upon existing search result clustering by integrating a NER system into their global system. This inclusion assigned particular importance to NEs, enhancing their clustering approach.

3. IMPLEMENTING THE NER AND THE TOOLS IT PROVIDES

This unique method of extracting information from text categorizes passage elements, identifying names, dates, places, and various other details. It is evident that a NER system successfully identifies five entities within the given frequent English and Arabic sentences, as illustrated in Figure 1.

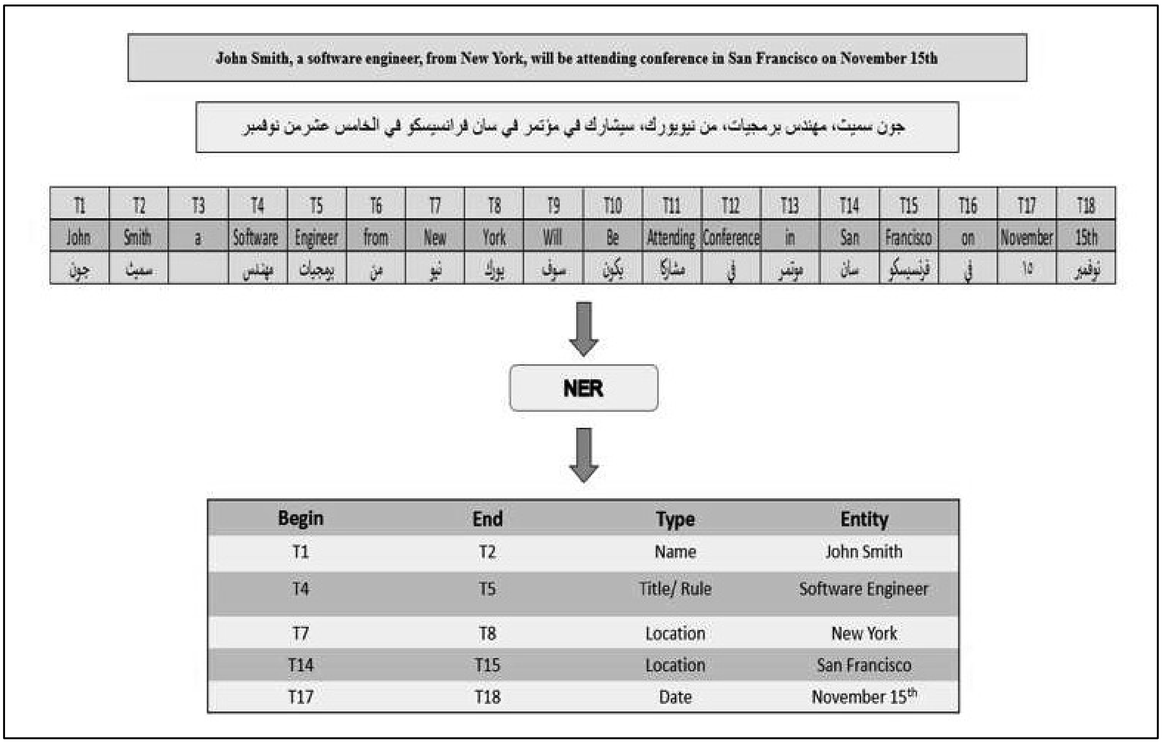

Figure 1. An illustration of the NER task after removing all punctuation. It is a typical example sentence and has been translated into Arabic as well.

As a sequence, the tools used to implement the NER will be explained in the following phase:

- **SpaCy:** Spacy is a Python package tailored for advanced NLP. It's open source, crafted for production environments, supports over 72 languages, offers 80 trained pipelines for nearly 24 languages, facilitates multitask learning with transformers like BERT, boasts state-of-the-art speed, and supports custom models in TensorFlow, PyTorch, and other libraries. Additionally, it includes a NER pipeline component known as an 'entity recognizer' [6].

- **Natural Language Tool Kit (NLTK):** The widespread Python infrastructure for working with human language data is NLTK. It offers text manipulation libraries for categorization, parsing, stemming, tagging, semantic reasoning, and tokenization. It includes an

NER module with a Maximum Entropy Classifier trained on the Automatic Content Extraction corpus [38]

- **Apache openNLP:** The Apache OpenNLP library, introduced by [38], is a machine learning toolset for natural language text processing. It supports various NLP tasks, including parsing, tokenization, part-of-speech tagging, and NER. The library incorporates Perceptron-based and Maximum Entropy machine learning. The Name Finder within OpenNLP can identify both NEs and numbers in text, utilizing specific models for recognition.

- **Tensorflow:** TensorFlow is an open-source math library created by Google. It's written in Python, C++, and CUDA and is designed for neural network models and machine learning. It works with numerical or one-hot-encoded data, not text, and is used in various applications, including Google Translate, text condensation, and NER [39].

- **Pytorch:** PyTorch, created by Facebook, is an open-source deep-learning package specifically crafted for Python. It serves as a prominent platform utilized in both industrial and educational contexts. PyTorch excels in tasks like image recognition and NLP, enabling various A.I. models' rapid and efficient construction to enhance scalability [40].

4. EXPLORING THE SIGNIFICANCE OF NER IN THE CONTEXT OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE

It is clear, according to our observation on the WOS index, that most NER research in recent years has shown its interest in the computer science Artificial intelligence field, as illustrated in Figure 2 and 3, and this is because the NER is the long-term and primary task of NLP and information extraction.

There are several reasons why NER research has increased over the years. The Developments in Deep Learning are achieved when NER systems have significantly elevated since the start of deep learning, especially with the advent of models such as transformer-based architectures like BERT, GPT, and their variations, and recurrent neural networks (RNNs). Researchers continuously investigate novel approaches to utilize these designs for enhanced precision and effectiveness in entity recognition [15]. Also, pre-trained language models use extensive corpora of textual data for training, and they have shown remarkable effectiveness on various NLP tasks, including NER. Scholars are investigating methods to optimize these trained models for particular domains or tasks, increasing the usability and efficacy of NER in a range of applications. Thus, the requirement of Growing Information Elicitation for Systems that can automatically obtain structured information from unstructured text is in greater demand as the amount of digital information keeps growing. By recognizing and classifying items in the text, NER is essential to information extraction and helps with applications like knowledge graph generation, search engines, and Q&A systems. Finally, Multilingual NER As information becomes more globally distributed, there is a growing need to create NER models that can efficiently handle different languages. Multilingual NER research aims to build models that recognize and categorize objects in text published in other languages. As a result, the advances in deep learning, the effect of already-trained language models, the growing need for information extraction, the availability of benchmark datasets and competitions that support research and development in this field are all responsible for the recent surge in NER research in computer science and artificial intelligence.

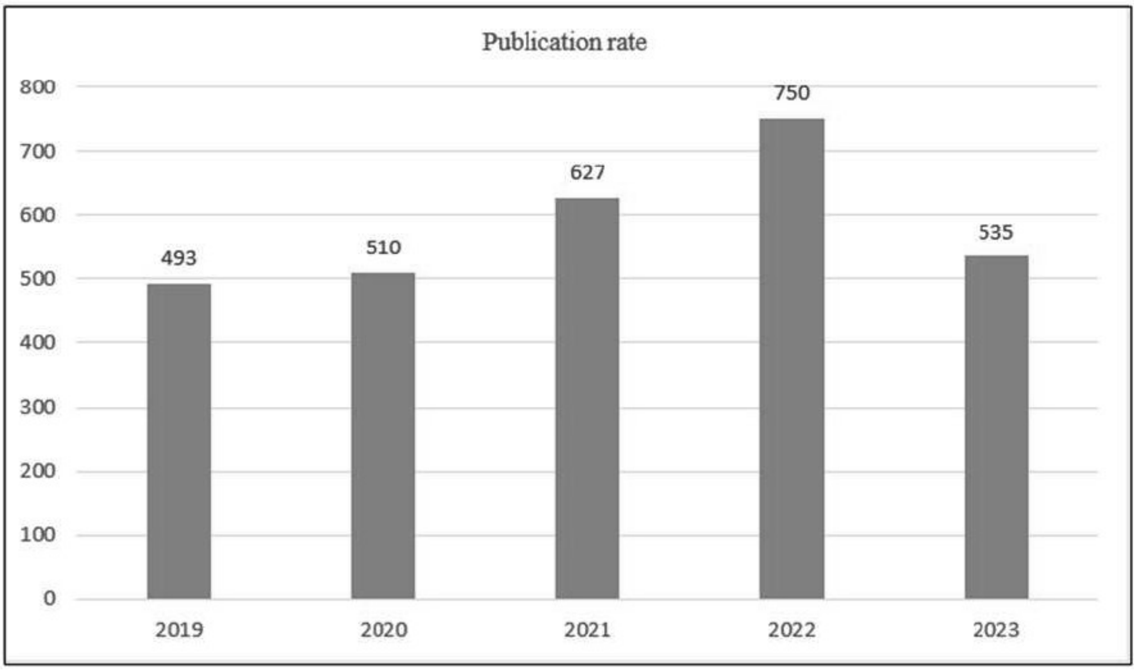

Figure 2. Illustrates the importance of NER computer science and AI

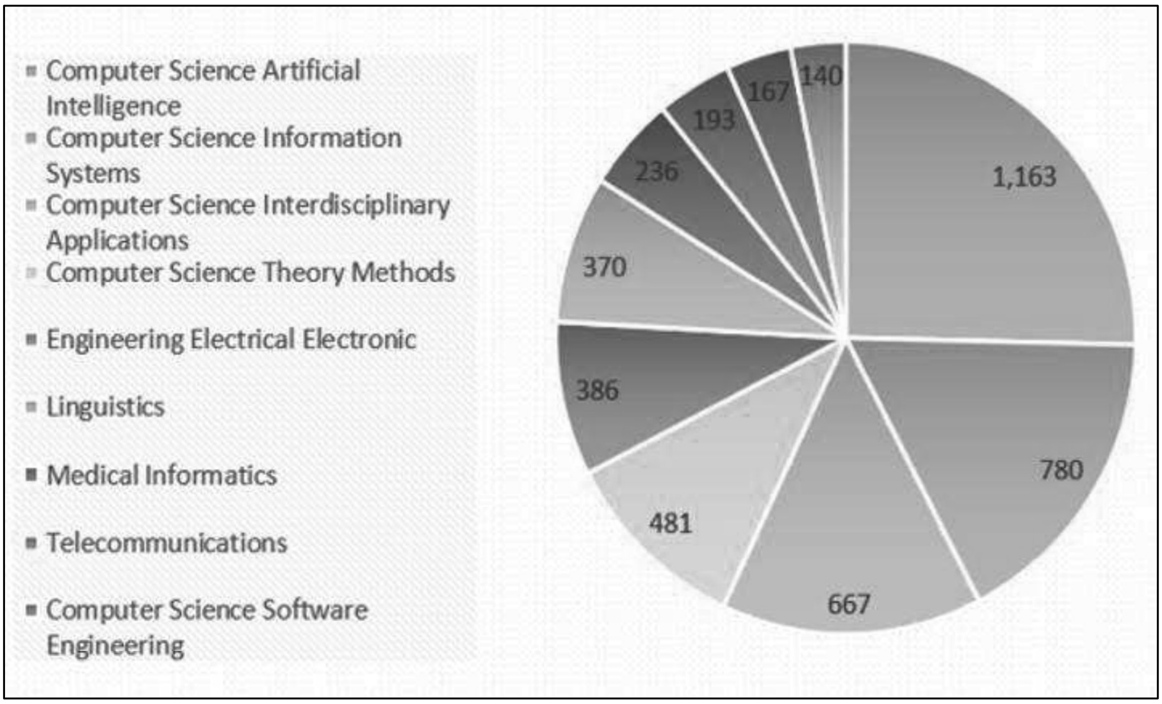

Figure 3. Publication rates of NER studies based on WoS indices

5. NER IN ENGLISH AND ARABIC

Specifically targeting English or Arabic, NER primarily focuses on extracting structured information from unstructured text to facilitate analysis and comprehension. In English NER (E-NER) or Arabic NER (A-NER), entities are typically organized into predefined groups like person, organization, location, date, time, money, percentage, etc. The process involves identifying and categorizing specific entities in English or Arabic text, presenting a structured representation for enhanced analysis, and understanding.

Figures 4. a and b illustrate the number of NER research studies conducted in Arabic and English over the past five years. It is evident that there is a significant volume of research in E-NER; on the other hand, research in A-NER is minimal, pointing to substantial opportunities for further exploration and advancements in the A-NER domain.

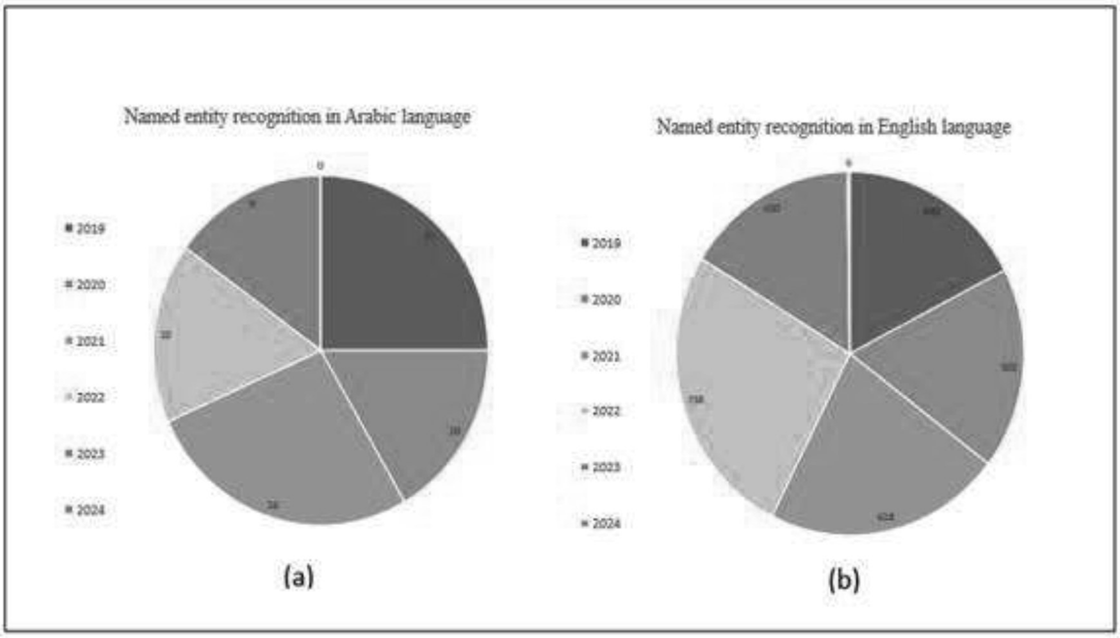

Figure 4. Quantitative Analysis of Research Papers on NER in English and Arabic Over the Past Five Years

5.1. Annual Publication Count

For both A-NER and E-NER research, a total of 60 and 2803 studies have been conducted between 2019 and 2024, according to the Web of Science (WOS) index. Figure 5 illustrates the annual evolution of these values. While ENER studies remain the most prevalent, there is a decline in the last data point for 2023. On the other hand, A-NER studies exhibit a scarcity of research in this area, including the latest data points. This scarcity highlights numerous opportunities for further study and research in the Arabic NER domain and the limitations and challenges that must be addressed. To provide insight into the preference for English studies, we present the ratio of English studies to Arabic studies in Figure 5. The figure indicates a relative increase in the preference for English studies compared to Arabic studies.

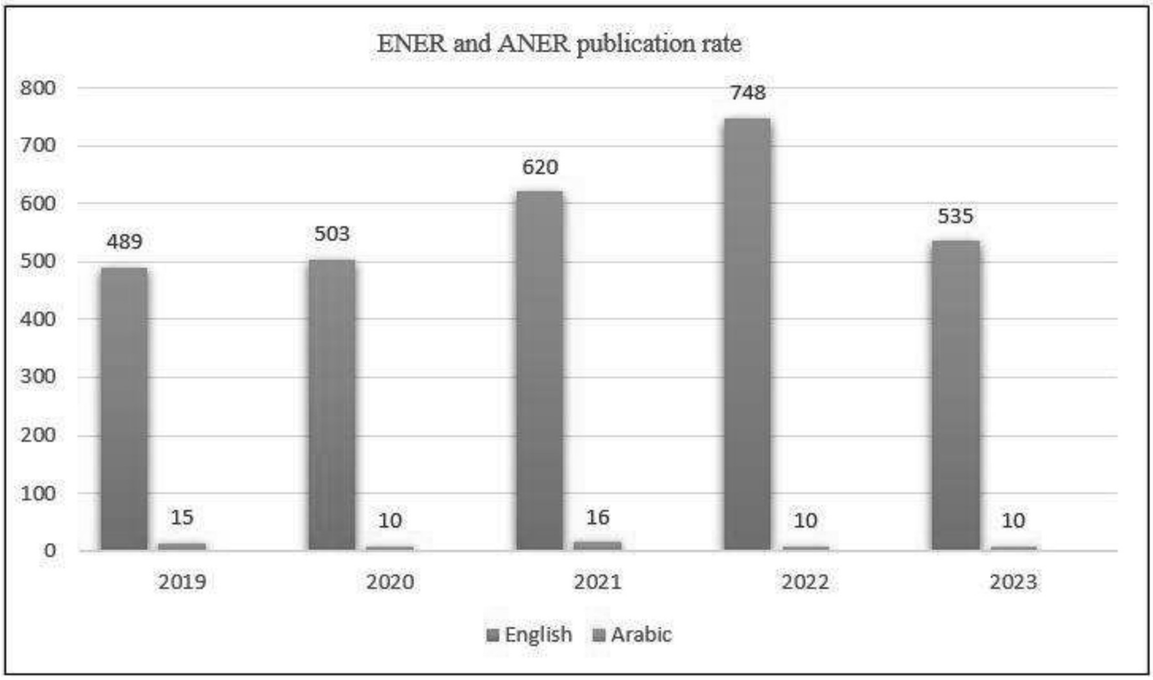

Figure 5. Papers on NER in Arabic and English languages in the last five years

6. LITERATURE REVIEW

In this study, the authors introduced a transfer learning and asymmetric tri-training leveraged by a biomedical NER model to address the scarcity of marked data in this domain. They achieved a remarkable 9%+ enhancement in exact F1 scores over the baseline BiLSTM-CRF model across four diverse datasets examined in this research [16].

Other researchers presented a detailed linguistic analysis of annotated data of isiXhosa NEs in this research, focusing on three primary NE types: person, location, and organization. They identified suffixes and capitalization as potential predictors for these NE types. They also introduce the NER and feature set evolved as part of the NCHLT version. While the recognizer demonstrates high accuracy (0.9713 overall), it accordingly exhibits lower retrieval (0.7409), particularly for human names (0.5963), resulting in a total F-score of 0.8406. This release significantly contributes to a historically under-resourced language, although there are opportunities for improving the NER. A labeled NER dataset for isiXhosa was made available through the NCHLT Text project [17].

An advanced deep learning model for nested NER is introduced in this analysis. Which captures interactive and contrasting span relationships. This method employs a scale transformation mechanism to include geometric information in span representations, enhancing span interaction modeling. Additionally, they employ a supervised contrastive loss to distinguish overlapping spans, achieving state-of-the-art or competitive results on three public nested NER datasets (ACE05-Chinese, ACE2005-English, and GENIA). For example, on ACE2005-Chinese, they achieved impressive precision, recall, and F1 scores of 87.11, 87.14, and 87.12, affirming their model's effectiveness [18].

The challenge of limited labeled data is addressed by introducing a method for cross-lingual NER using an attention mechanism and adversarial training. It employs resource-rich language data to enhance recognition in low-resource languages, achieving a significant F1 score improvement of 6.29% compared to the baseline in an English-Chinese NER task. The method effectively tackles the problem of long-sequence semantic dilution and shows promise in improving NER results in low-resource languages. This study employs Conll2003 datasets for training and conducts experiments using WeiboNER and People's Daily News datasets [19].

Additionally, a novel method for Chinese NER focuses on addressing fuzzy lexical boundaries. This approach employs multi-head attention and combines Word2vec, HMM, and ALBERT for word, boundary, and character vector extraction, respectively. These vectors are fused using a Feedforward-attention mechanism and further refined by BiL-LSTM. Multi-head attention is used to extract crucial word information from text features, followed by text label classification using conditional random field (CRF). The model exhibits impressive accuracy, achieving a rate of 95.81% in tests on WeiboNER, MSRA, and CLUENER 2020 datasets [20].

The authors [21] investigated the optimal representation of NE tags within a morpheme-based system. They developed an algorithm to convert Korean corpora containing NEs from word-based and syllable-based formats to the proposed morpheme-based format. Using both traditional and neural models, they assessed the effectiveness of this new format, showing its viability. They also demonstrated how the models' performance varied with the inclusion of language-specific features and external conditions. This evaluation included different data types, such as the original segmentation and various tagging formats. The Korean NER data presented in this study is sourced from NAVER, initially compiled for a Korean NER competition held in 2018.

Methods for generating prompts in NER are discussed in this article, treating recognition as answering questions about NEs. It examines the impact of prompts on nested NER in a few-shot setting. Various prompt-based approaches were tested using RuNNE-2022 data. In the general task, multiple approaches showed similar results, with the best performance achieved by combining prompts. In the few-shot task, the most effective prompts were full lexical and full marker prompts, accounting for nested NEs. The NEREL dataset WAS USED [22].

Within the confines of this study, it introduces a novel approach for social media NER using data augmentation. It begins by pre-training a language model with BERT to create semantic word vectors based on contextual information. Then, it augments the data by finding similar entities and applying substitution or semantic transformations. The Bi-LSTM model processes the input, followed by fusion and fine-tuning for the best label. Additionally, the self-attentive layer captures critical features, reducing the need for external information. Experimental results on WNUT16, WNUT17, and OntoNotes 5.0 datasets validate the model's effectiveness with NEREL dataset utilization [23].

This study [24] investigates Telugu NER using various word embeddings, including Glove, FastText, Word2Vec, Contextual String embedding, and BERT embeddings generated from Telugu Wikipedia articles. Deep learning models were constructed using these embeddings as input. Additionally, they examined the impact of incorporating handmade features into the word embeddings. The experimental results reveal that BERT embeddings with custom-made features Super passed other models, achieving an F1-Score of 96.32% on the FIRE-2018 benchmark dataset.

This essay [25] offers five custom-built NER models and evaluates their performance towards three widely used pre-trained models for extracting location names. They assess these models using manually annotated Wikipedia articles, measuring their performance with the F1 score pattern. Their best model conquered an impressive F1 score of 0.939, outperforming the top pre-built model, which scored 0.730. They apply their model to extract location names from Wikipedia topics in Great Britain, showcasing its capability to preciously identify mysterious location names from user-contributed online geographic data. Wikipedia offers a well-structured text with a consistent writing style and minimal misspellings, making it an excellent data source for our research, which accessed Wikipedia text data through DBpedia.

This paper [14] addresses the challenge of NER in user-generated texts from sources like social media, which often contain unconventional words and abbreviations. The authors introduce two methods, alias augmentation, and typo augmentation, to generate weakly labeled data from unlabeled texts like Wikipedia and Tweets. These methods, combined with transfer learning, significantly enhance NER performance. Their experimental results show that their approach achieves the highest reported F1-score of 51.43% in NER.

In this study [26], the authors have taken the initiative to create a relatively large dataset for Amharic NER and have made it publicly accessible. They've used this newly developed dataset to construct multiple systems for recognizing NEs in Amharic text. These systems are built on cutting-edge deep learning techniques, containing transfer learning with BLSTM and RoBERTa memory attached with a conditional accidental fields layer. To address the issue of imbalanced classification, the authors have applied the Synthetic Minority Over-sampling Technique. The most successful system, which is based on RoBERTa, has achieved an impressive F1-score of 93%. This remarkable achievement represents a new work of art in Amharic NER.

In this research [27], the investigators introduce the BERT-Span model for the purpose of performing NER in the domain of Chinese rehabilitation medicine. They began by gathering rehabilitation-related data from various sources to create a specialized corpus in the field of rehabilitation medicine. Subsequently, they fine-tuned the Bidirectional Encoder Representation from Transformers (BERT) using this newly created rehabilitation medicine corpus. Within the context of the rehabilitation medicine corpus, the authors utilize BERT to extract feature vectors corresponding to rehabilitation medicine entities within the text. The span model then employs these vectors to perform the annotation of rehabilitation medicine entities. Impressively, this approach yields an F1 score of 93.07%, indicating high precision and recall in identifying these entities.

The author introduces two noise reduction models [28], Shared Labels and Dynamic Splicing, built upon XLNet encoding, a permutation-based language pre-training model, and employs Conditional Random Field (CRF) for decoding. Through rigorous evaluation of 15 biomedical NER datasets, these models have exhibited remarkable improvements in the average F1-score, with gains of 1.504 and 1.48, respectively. They have set new performance standards by achieving state-of-the-art results on seven evaluated datasets. Furthermore, in-depth analysis confirms the effectiveness of these models in enhancing CRF-based recognition. The author also provides insights into the suitability of these models across diverse data characteristics.

Other researchers [29] introduced a deep learning model that fine-tunes the BERT model for A-NER and classification. They deployed pre-trained BERT context insertion as input features for a Bidirectional Gated Recurrent Unit (BGRU). They were fine-tuned with two annotated A-NER datasets. Experimental findings reveal that the proposed model outperforms existing cutting-edge A-NER models, achieving F-measure values of 92.28% on the A-NERCorp dataset and 90.68% on the combined A-NERCorp and AQMAR datasets.

The researchers [30] explore two approaches for extracting dataset mentions from scientific papers. The first involves a two-step process, using a binary classifier followed by transformer models for extraction. The second is a one-step approach using transformers in question-answering mode. While the two-step method extracts more mentions, it has a lower F1-score of 62.7%. In contrast, the one-step approach with DeBERTa achieved a high F1-score of 92.88% but may miss some mentions. This study is based on the Coleridge Initiative 'Show US the Data' dataset containing 14.3k scientific papers with around 35k dataset mentions.

This research [31] achieved a multitask processing model, BilSTM-BERT-AM-CRF, which leverages BERT is used to extract dynamic word vectors enriched with contextual information. After further training through the BiLSTM module, the results are fed into a CRF layer for decoding. The model used the attention mechanism network to learn from two Chinese datasets. Subsequently, the CRF is employed to classify and extract the observation annotation sequence, yielding the results. Compared to several previous single-task models, this multitasks model has exhibited significant F1 score improvements in both MASR and People's Daily datasets, with gains of 0.55% and 3.41%, underscoring the effectiveness of multitask learning in Chinese NER.

Specific tasks are addressed in this research as follows [32]: (1) It introduces a data transfer method based on entity features. The task involves electing representative entity features by measuring the similarity of feature distribution between low and high-resource data. This method calculates the distribution gap between the two domains and trains the model using high-resource data. (2) It proposes an entity boundary detection method that utilizes BiLSTM+CRF as the core structure, incorporating character boundary information to enhance entity boundary recognition through an attention network. (3) The research compares these methods with multiple NER baseline methods in experiments across various datasets. The model's results suggest that the model presented in this study enhances NER accuracy by 1%, F1 value by 2%, and recall rate by 2% on average in domains with limited resources.

The researchers [33] proposed a novel method for recognizing NEs using controlled attention. They embed task-related cues in sentences to mark possible entity boundaries, then process these sentences with a deep network to learn discriminative entity-relevant representations. In experiments on English and Chinese corpora, their method outperforms existing models, particularly in nested NER. It offers three key advantages: enhancing the neural network's awareness of entity boundaries, enabling a focus on task-related semantic features, and potential extension to other NLP tasks like entity relation and event extraction. This paper [34] aimed to introduce a RoBERTa-GlobalPointer approach for legal document NER. It has combined character and word-level features to capture entity context. GlobalPointer calculates entity scores, and a balanced softmax function determines entity types. Evaluation of a Chinese judicial dataset demonstrates its superiority over current techniques.

Finally, a new approach [35] has been proposed to represent categories using multiple prototypes in a hybrid manner so that labels for entity categories are included by inserting posters after entities into support sentences. Entity-level prototypes are formed by averaging contextual token embeddings, and label-level prototypes are established from contextual label embeddings. For non-entity classes, token embeddings represent them, constituting the multi-prototype. This hybrid strategy improved class representations from support examples, achieving a 3%~10% increase in F1 scores over prior models in a rigorous few-shot NER setting.

Deep learning has been a potent technique for directly learning feature representations from data in recent years, and it has produced tremendous advances in the field of natural NLP. When applied in NER, deep learning can acquire sophisticated hidden representations without extensive domain knowledge or complex feature engineering. Thus, deep learning-based techniques in NER have vastly outperformed conventional rule-based and statistical-based techniques. In Table 1 collection of the 20 most recent research initiatives in NER summarizes highly effective methodologies for identifying NE's across various languages and domains, imparting a thorough comprehension of state-of-the-art techniques. By incorporating this information, researchers can better understand the current methodological scenario, enabling them to develop and implement promising approaches in their studies.

Table 1. A comprehensive overview of recent advances in NER over the last five years

Ref. / Year	Language	Method	Dataset	Result
[39] 2020	Chinese	Extraction of multi-scale local context features using a CNN with varying kernel sizes. The specific model used is JMCA-ADP.	AgCNER	P=93.67 R= 95.14 F1= 94.15
[40] 2022	Arabic	Build the BERT Bi-LSTM-CRF model.	Modern Standard Arabic (MSA)	67.40
[41] 2021		A cutting-edge model that combines the power of deep neural networks using a bidirectional Long Short-Term Memory (LSTM) approach along with CNN	CoNLL 2003	
[42] 2022	Turkish	obtained by Transformer-based language models	1-WikiANN-tr 2-News Articles 3-ATISNER 4-TWNER 5-WikiANN-tr 6- FBNER	From 80.8 To 96.1
[43] 2020	Korean	Bi-LSTM-CRF NER Tagger	1-Klpexpo 2016 NER 2- Naver NLP	1'st dataset 86.27 2'nd dataset 91.07
[44]	English	BERT, ELMO and BiLSTM-	1- CoNLL-	1'st dataset

2021		CRF (BLC) model	2003 English NER 2- Onto Notes 5.0	95.57 2'nd dataset 89.21
[45] 2023	Chinese	USAF: the textual feature extractor, the acoustic feature synthesizer, and the multimodal interaction module.	1- CNERTA 2- Aishell3-NER 3- MSRA	1'st dataset 78.16 2'nd dataset 93 3'rd dataset 94.74
[46] 2023	Shahmuki	CNN-BiLSTM-CRF + W2V + ELMo	EMILLE and Wikipedia	98.46
[47] 2023	English and Spanish	Conditional Random Fields (CRFs) and Bi-LSTM	DANN	In Spanish 84.6 In English 0.861
[48] 2022	French	CAS Privacy-Preserving Mimic Model	Private	70.6
[49] 2022	________	BiLSTM and dilated convolutional neural network (DCNN)	1- Hospital-BJ 2- CCKS-2017 3-CCKS-2018	1'st dataset 92.85 2'nd dataset 98.52 3'rd dataset 93.79
[50] 2022	________	LELNER model	1-BC5CDR 2-CONLL 3- SemEval-lap	1'st dataset 75.52(±0.2) 2'nd dataseT 87.12(±0.3) 3'rd dataset 72.41(±0.3)
[51] 2021	________	unsupervised cross-domain model	1-CoNLL2003 2-Twitter 3-SciTech	supervised using 40% target with Unsupervised= 0.13% F1
[52] 2023	Chinese	Multiple features and the geological pre-training model (GeoBERT)	GNER	79.60
[53] 2022	________	BERT-BiLSTM-CRF	1-CoNLL2003 2-OntoNotes-5.0	72.55

			3- WNUT2016	
[54] 2020	________	The DBNER method employs LSTM and CRF decoding models for recognizing bug-specific entities.	Bugzilla	91.19%.
[55] 2023	Arabic (Algerian)	1. Data gathering 2. Initial processing 3. The annotation procedure	DzNER	------------
[56] 2020	________	BLSTM combined with conditional random field (CRF) and dynamic RNN)	1- BioCreAtIvE II GM 2- JNLPBA 3- NCBI	1'st dataset BioCreAtIvE 89.98 2'nd dataset NCBI 90.84
[57] 2022	________	span-based nested NER model with BERT	1- GENIA 2- ACE2004 3- ACE2005	1'st dataset 79.46% 2'nd dataset 87.30% 3'rd dataset 85.24
[58] 2022	English and Greek	Transformer-based networks and BLSTM	1-ATIS-EN (English) 2-ATIS-GR (Greek)	1'st dataset 96.4 2'nd dataset 95.7

7. DISCUSSION AND CONCLUSION

NER for clinical tasks in non-English languages is challenged by limited data. Solutions, such as cross-lingual transfer (CLT) using multilingual models or translation-based methods without requiring annotated data [59], are employed. However, the challenge remains in developing NER systems that can perform effectively with limited marked information, particularly in languages with fewer resources [60]. As exemplified by Wikipedia, the ongoing challenge lies in distinguishing between entities with similar names and linking them to unique identifiers in knowledge bases [61]. Recent growth in NLP has been experienced across various research domains, driven by continuous advancements in big data, deep learning, and machine learning. The promise of enhanced efficiency and accuracy in a wide range of tasks characterizes the future of NLP. A comprehensive overview of the latest breakthroughs in NER, including cutting-edge techniques, algorithms, and tools shaping the dynamic landscape of NER in the field of NLP, is aimed to be provided in our paper. These developments are explored to offer a thorough understanding of NER and its practical implications. The crucial role played by NER in the fields of Computer Science and Artificial Intelligence is essential to emphasize. Acting as a linchpin, NER facilitates the navigation of the complex web of data and language, enabling a deeper understanding and more nuanced interaction with information. The extraction of entities from text by NER technology enhances the contextual understanding of language, contributing significantly to the overall advancement of these fields.

Our exploration aims to present the current state of NER and highlight its significance in computational advancements. The interplay between NLP, big data, and advanced learning methodologies promises a future where language comprehension and interaction reach unprecedented heights.

REFERENCES

[1] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," IEEE Computational Intelligence Magazine, vol. 9, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 48–57, 2014. doi: 10.1109/MCI.2014.2307227.
[2] Y. Kang, Z. Cai, C. W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," Journal of Management Analytics, vol. 7, no. 2. Taylor and Francis Ltd., pp. 139–172, Apr. 02, 2020. doi: 10.1080/23270012.2020.1756939.
[3] R. Grishman, "Message Understanding Conference-6: A Brief History."
[4] Y. Wang, H. Tong, Z. Zhu, and Y. Li, "Nested Named Entity Recognition: A Survey," ACM Trans Knowl Discov Data, vol. 16, no. 6, Jul. 2022, doi: 10.1145/3522593.
[5] Y. Kang, Z. Cai, C. W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," Journal of Management Analytics, vol. 7, no. 2. Taylor and Francis Ltd., pp. 139–172, Apr. 02, 2020. doi: 10.1080/23270012.2020.1756939.
[6] S. Vychegzhanin and E. Kotelnikov, "Comparison of named entity recognition tools applied to news articles," in Proceedings - 2019 Ivannikov Ispras Open Conference, ISPRAS 2019, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 72–77. doi: 10.1109/ISPRAS47671.2019.00017.
[7] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification." [Online]. Available: http://projects.ldc.upenn.edu/gale/
[8] H. Fei, Y. Ren, and D. Ji, "Dispatched attention with multitask learning for nested mention recognition," Inf Sci (N Y), vol. 513, pp. 241–251, Mar. 2020, doi: 10.1016/j.ins.2019.10.065.
[9] J. M. Giorgi and G. D. Bader, "Towards reliable named entity recognition in the biomedical domain," Bioinformatics, vol. 36, no. 1, pp. 280–286, Jan. 2020, doi: 10.1093/bioinformatics/btz504.
[10] L. Gligic, A. Kormilitzin, P. Goldberg, and A. Nevado-Holgado, "Named entity recognition in electronic health records using transfer learning bootstrapped Neural Networks," Neural Networks, vol. 121, pp. 132–139, Jan. 2020, doi: 10.1016/j.neunet.2019.08.032.
[11] E. Aramaki, Y. Miura, M. Tonoike, and F. Xerox, "TEXT2TABLE: Medical Text Summarization System based on Named Entity Recognition and Modality Identification Tomoko OHKUMA Hiroshi MASHUICHI Kazuhiko OHE." [Online]. Available: http://www.seas.upenn.edu/~pdtb/
[12] M. Khodabakhsh, M. Kahani, E. Bagheri, and Z. Noorian, "Detecting life events from twitter based on temporal semantic features," Knowl Based Syst, vol. 148, pp. 1–16, May 2018, doi: 10.1016/j.knosys.2018.02.021.
[13] A. Ritter, S. Clark, and O. Etzioni, "Named Entity Recognition in Tweets: An Experimental Study," Association for Computational Linguistics.

[14] J. Kim, Y. Kim, and S. Kang, "Weakly labeled data augmentation for social media named entity recognition," Expert Syst Appl, vol. 209, Dec. 2022, doi: 10.1016/j.eswa.2022.118217.
[15] M. Nazmus Saadat and M. Shuaib, "Advancements in Deep Learning Theory and Applications: Perspective in 2020 and beyond," in Advances and Applications in Deep Learning, IntechOpen, 2020. doi: 10.5772/intechopen.92271.
[16] M. Bhattacharya, S. Bhat, S. Tripathy, A. Bansal, and M. Choudhary, "Improving biomedical named entity recognition through transfer learning and asymmetric tri-training," in Procedia Computer Science, Elsevier B.V., 2022, pp. 2723–2733. doi: 10.1016/j.procs.2023.01.244.
[17] R. Eiselen and A. Bukula, "IsiXhosa Named Entity Recognition Resources," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 2, Dec. 2022, doi: 10.1145/3531478.
[18] Q. Zheng et al., "Exploring Interactive and Contrastive Relations for Nested Named Entity Recognition," IEEE/ACM Trans Audio Speech Lang Process, vol. 31, pp. 2899–2909, 2023, doi: 10.1109/TASLP.2023.3293047.
[19] H. Wang, L. Zhou, J. Duan, and L. He, "Cross-Lingual Named Entity Recognition Based on Attention and Adversarial Training," Applied Sciences (Switzerland), vol. 13, no. 4, Feb. 2023, doi: 10.3390/app13042548.
[20] Z. Sun and X. Li, "Named Entity Recognition Model Based on Feature Fusion," Information (Switzerland), vol. 14, no. 2, Feb. 2023, doi: 10.3390/info14020133.
[21] Y. Chen, K. T. Lim, and J. Park, "Korean named entity recognition based on language-specific features," Nat Lang Eng, 2023, doi: 10.1017/S1351324923000311.
[22] I. S. Rozhkov and N. V. Loukachevitch, "Prompts in Few-Shot Named Entity Recognition," Pattern Recognition and Image Analysis, vol. 33, no. 2, pp. 122–131, Jun. 2023, doi: 10.1134/S1054661823020104.
[23] W. Liu and X. Cui, "Improving Named Entity Recognition for Social Media with Data Augmentation," Applied Sciences (Switzerland), vol. 13, no. 9, May 2023, doi: 10.3390/app13095360.
[24] S. K. Gorla, S. S. Tangeda, L. B. M. Neti, and A. Malapati, "Telugu named entity recognition using bert," Int J Data Sci Anal, vol. 14, no. 2, pp. 127–140, Aug. 2022, doi: 10.1007/s41060-021-00305-w.
[25] C. Berragan, A. Singleton, A. Calafiore, and J. Morley, "Transformer based named entity recognition for place name extraction from unstructured text," International Journal of Geographical Information Science, vol. 37, no. 4, pp. 747–766, 2023, doi: 10.1080/13658816.2022.2133125.
[26] E. C. Jibril and A. C. Tantug, "ANEC: An Amharic Named Entity Corpus and Transformer Based Recognizer," IEEE Access, vol. 11, pp. 15799–15815, 2023, doi: 10.1109/ACCESS.2023.3243468.
[27] J. Zhong, Z. Xuan, K. Wang, and Z. Cheng, "A BERT-Span model for Chinese named entity recognition in rehabilitation medicine," PeerJ Comput Sci, vol. 9, 2023, doi: 10.7717/peerj-cs.1535.
[28] Z. Chai et al., “Noise Reduction Learning Based on XLNet-CRF for Biomedical Named Entity Recognition,” IEEE/ACM Trans Comput Biol Bioinform, vol. 20, no. 1, pp. 595–605, Jan. 2023, doi: 10.1109/TCBB.2022.3157630.
[29] N. Alsaaran and M. Alrabiah, "Arabic Named Entity Recognition: A BERT-BGRU Approach," Computers, Materials and Continua, vol. 68, no. 1, pp. 471–485, Mar. 2021, doi: 10.32604/cmc.2021.016054.
[30] Y. Younes and A. Scherp, "Question Answering versus Named Entity Recognition for Extracting Unknown Datasets," IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3309148.

[31] X. Tang, Y. Huang, M. Xia, and C. Long, "A Multitask BERT-BiLSTM-AM-CRF Strategy for Chinese Named Entity Recognition," Neural Process Lett, vol. 55, no. 2, pp. 1209–1229, Apr. 2023, doi: 10.1007/s11063-022-10933-3.
[32] H. Kang, J. Xiao, Y. Zhang, L. Zhang, X. Zhao, and T. Feng, "A Research Toward Chinese Named Entity Recognition Based on Transfer Learning," International Journal of Computational Intelligence Systems, vol. 16, no. 1, Dec. 2023, doi: 10.1007/s44196-023-00244-3.
[33] Y. Chen, R. Huang, L. Pan, R. Huang, Q. Zheng, and P. Chen, "A Controlled Attention for Nested Named Entity Recognition," Cognit Comput, vol. 15, no. 1, pp. 132–145, Jan. 2023, doi: 10.1007/s12559-023-10112-z.
[34] X. Zhang, X. Luo, and J. Wu, "A RoBERTa-GlobalPointer-Based Method for Named Entity Recognition of Legal Documents," in Proceedings of the International Joint Conference on Neural Networks, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IJCNN54540.2023.10191275.
[35] P. Liu, Y. Guo, F. Wang, and G. Li, "Chinese named entity recognition: The state of the art," Neurocomputing, vol. 473, pp. 37–53, Feb. 2022, doi: 10.1016/j.neucom.2021.10.101.
[36] B. Jehangir, S. Radhakrishnan, and R. Agarwal, "A survey on Named Entity Recognition — datasets, tools, and methodologies," Natural Language Processing Journal, vol. 3, p. 100017, Jun. 2023, doi: 10.1016/j.nlp.2023.100017.
[37] USENIX Association., ACM SIGMOBILE., ACM Special Interest Group in Operating Systems., and ACM Digital Library., Papers presented at the Workshop on Wireless Traffic Measurements and Modeling : June 5, 2005, Seattle, WA, USA. USENIX Association, 2005.
[38] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2019.
[39] X. Guo et al., "Chinese agricultural diseases and pests named entity recognition with multi-scale local context features and self-attention mechanism," Comput Electron Agric, vol. 179, Dec. 2020, doi: 10.1016/j.compag.2020.105830.
[40] B. A. Benali, S. Mihi, N. Laachfoubi, and A. A. Mlouk, "Arabic Named Entity Recognition in Arabic Tweets Using BERT-based Models," in Procedia Computer Science, Elsevier B.V., 2022, pp. 733–738. doi: 10.1016/j.procs.2022.07.109.
[41] C. Suman, S. M. Reddy, S. Saha, and P. Bhattacharyya, "Why pay more? A simple and efficient named entity recognition system for tweets," Expert Syst Appl, vol. 167, Apr. 2021, doi: 10.1016/j.eswa.2020.114101.
[42] O. Ozcelik and C. Toraman, "Named entity recognition in Turkish: A comparative study with detailed error analysis," Inf Process Manag, vol. 59, no. 6, Nov. 2022, doi: 10.1016/j.ipm.2022.103065.
[43] G. Jin and Z. Yu, "A Korean named entity recognition method using Bi-LSTM-CRF and masked self-attention," Comput Speech Lang, vol. 65, Jan. 2021, doi: 10.1016/j.csl.2020.101134.
[44] M. Affi and C. Latiri, "BE-BLC: BERT-ELMO-based deep neural network architecture for English named entity recognition task," in Procedia Computer Science, Elsevier B.V., 2021, pp. 168–181. doi: 10.1016/j.procs.2021.08.018.
[45] Y. Liu, S. Huang, R. Li, N. Yan, and Z. Du, "USAF: Multimodal Chinese named entity recognition using synthesized acoustic features," Inf Process Manag, vol. 60, no. 3, May 2023, doi: 10.1016/j.ipm.2023.103290.
[46] A. Tehseen, T. Ehsan, H. Bin Liaqat, X. Kong, A. Ali, and A. Al-Fuqaha, "Shahmukhi named entity recognition by using contextualized word embeddings," Expert Syst Appl, vol. 229, Nov. 2023, doi: 10.1016/j.eswa.2023.120489.

[47] H. Fabregat, A. Duque, J. Martinez-Romo, and L. Araujo, "Negation-based transfer learning for improving biomedical Named Entity Recognition and Relation Extraction," J Biomed Inform, vol. 138, Feb. 2023, doi: 10.1016/j.jbi.2022.104279.

[48] N. Bannour, P. Wajsbürt, B. Rance, X. Tannier, and A. Névéol, "Privacy-preserving mimic models for clinical named entity recognition in French," J Biomed Inform, vol. 130, Jun. 2022, doi: 10.1016/j.jbi.2022.104073.

[49] R. Zhang, P. Zhao, W. Guo, R. Wang, and W. Lu, "Medical named entity recognition based on dilated convolutional neural network," Cognitive Robotics, vol. 2, pp. 13–20, Jan. 2022, doi: 10.1016/j.cogr.2021.11.002.

[50] Z. Zhang, H. Zhang, Q. Wan, and J. Liu, "LELNER: A Lightweight and Effective Low-resource Named Entity Recognition model," Knowl Based Syst, vol. 251, Sep. 2022, doi: 10.1016/j.knosys.2022.109178.

[51] Q. Peng, C. Zheng, Y. Cai, T. Wang, H. Xie, and Q. Li, "Unsupervised cross-domain named entity recognition using entity-aware adversarial training," Neural Networks, vol. 138, pp. 68–77, Jun. 2021, doi: 10.1016/j.neunet.2020.12.027.

[52] Q. Qiu et al., "Chinese engineering geological named entity recognition by fusing multi-features and data enhancement using deep learning ☆," Expert Syst Appl, vol. 238, pp. 957–4174, 2024, doi: 10.24433/CO.6163056.v1.

[53] W. Li et al., "UD_BBC: Named entity recognition in social network combined BERT-BiLSTM-CRF with active learning," Eng Appl Artif Intell, Nov. 2022, doi: 10.1016/j.engappai.2022.105460.

[54] C. Zhou, B. Li, and X. Sun, "Improving software bug-specific named entity recognition with deep neural network," Journal of Systems and Software, vol. 165, Jul. 2020, doi: 10.1016/j.jss.2020.110572.

[55] A. H. Dahou and M. A. Cheragui, "DzNER: A large Algerian Named Entity Recognition dataset," Natural Language Processing Journal, vol. 3, p. 100005, Jun. 2023, doi: 10.1016/j.nlp.2023.100005.

[56] S. Gajendran, M. D, and V. Sugumaran, "Character level and word level embedding with bidirectional LSTM – Dynamic recurrent neural network for biomedical named entity recognition from literature," J Biomed Inform, vol. 112, Dec. 2020, doi: 10.1016/j.jbi.2020.103609.

[57] C. Liu, H. Fan, and J. Liu, "Handling negative samples problems in span-based nested named entity recognition," Neurocomputing, vol. 505, pp. 353–361, Sep. 2022, doi: 10.1016/j.neucom.2022.07.012.

[58] S. Rizou, A. Paflioti, A. Theofilatos, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Multilingual Name Entity Recognition and Intent Classification employing Deep Learning architectures," Simul Model Pract Theory, vol. 120, Nov. 2022, doi: 10.1016/j.simpat.2022.102620.

[59] F. Gaschi, X. Fontaine, P. Rastin, and Y. Toussaint, "Multilingual Clinical NER: Translation or Cross-lingual Transfer?," 2023.

[60] Y. Hou et al., "Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.05702

[61] L. Logeswaran, M.-W. Chang, K. Lee, K. Toutanova, J. Devlin, and H. Lee, "Zero-Shot Entity Linking by Reading Entity Descriptions," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.07348